DOI:
10.1021/acs.jpcb.6b12542

Document Version
Peer reviewed version

Salt Interactions in Solution Prevent Urea from Direct Association with a Peptide Backbone

Nicola Steinke,[†] Anna Genina,[‡] Christian D. Lorenz,[*,‡] and Sylvia E. McLain[*,†]

Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK, and Department of Physics, King's College London, London SE1 9NH, UK

E-mail: chris.lorenz@kcl.ac.uk; sylvia.mclain@bioch.ox.ac.uk

[*]To whom correspondence should be addressed
[†]University of Oxford
[‡]King's College London

Abstract

There is an ongoing debate as to how urea denatures proteins in solution. Using a combination of neutron scattering and computer simulation of a model peptide KGPGK, it was found that the ionic strength and pH have a significant impact on the urea-peptide interaction. From the work presented here, it appears that urea first and foremost decreases the charge-based interactions in solution, such as the TFA-TFA association, before interacting with the peptide backbone via hydrogen bonds. This gives insight into the pH and salt concentration dependency of urea-caused protein denaturation and might unify direct and indirect theories of urea-induced protein denaturation. The observed differences between MD and neutron and X-ray diffraction data might show that MD, in this particular case, underestimates the influence of charged fluorinated solutes.

Introduction

How and why small molecules, such as urea, induce the denaturation of proteins in aqueous solutions is still not well understood. There are many different theories as to how urea acts to cause unfolding.[1,2] One of the most prevalent proposes that urea interacts directly[3] with the protein side chains[4–6] or backbone[5,7–10] to initiate the unfolding process. It is debated whether or not these direct urea-protein interactions are mainly via hydrophobic residues[11] or hydrophilic in nature.[8] In opposition to this direct urea-peptide interaction, it has also been proposed that urea acts indirectly by changing the bulk water-water interactions, leading to a decrease in protein hydration, which destabilizes the protein, causing it to unfold[3] and expose hydrophobic parts of the protein.[12]

Previous studies have shown an ambiguous influence of salt on the ability of urea to denature proteins. Lower salt concentrations are thought to stabilize against urea-induced denaturation[13,14] whereas higher salt concentrations appear to destabilize the protein in the presence of urea.[13] In addition, osmolytes such as trimethylamine N-oxide (TMAO)

and proline are known to protect against urea-induced denaturation,[15] and certain proteins from halophilic organisms are resistant against urea denaturation in the presence of high salt concentrations.[16] Recently, the protective properties of osmolytes was shown by NMR to be caused by a direct hydrogen bonding interaction between the osmolyte betaine and urea.[17] In other conditions, a low[18] or high pH,[19] has been shown to facilitate urea-induced denaturation.

Interestingly, although it is well-known that proteins only function under very specific conditions, many *in vitro* investigations of proteins, such as X-ray crystallography and other biochemical assays, use high-salt buffers and unphysiological pH conditions in order to stabilize proteins. There is clearly a conundrum which is not easy to overcome in biophysical investigations of proteins; while *in vivo* proteins only function in their native form under heavily prescribed conditions, unphysiological conditions must be used in order to perform studies on proteins *in vitro*.

In order to assess the balance between salts and urea, and how they interact with different parts of a protein in solution, here the atomic scale details of a model peptide - lysine-glycine-proline-glycine-lysinamide (KGPGK; Fig. 1) - both with and without urea in solution have been investigated. In both solutions, both KGPGK and the anion trifluoroacetic acid (TFA) are present, where TFA is a commonly used counter anion in peptide synthesis and is therefore present in many biophysical investigations of peptides.

KGPGK contains a sequence which commonly forms turns in in many proteins (XPGX'),[20–25] and allows for the observation of a peptide in a variety of conformations, so that the role of water and urea can be probed directly. The XGPGX' sequence is found 1347 times in the UniProt Protein Database,[26] mostly in loop-sequences; 39% in β-turns and 10 % in γ-turns. The lysine residues were chosen in order to add both more solubility and flexibility to the system.

By using a combination of neutron diffraction with isotopic substitution and X-ray diffraction with computational techniques (Molecular Dynamics (MD) and Empirical Potential

Structure Refinement (EPSR;[27] a Monte Carlo-based simulation that is constrained by a set of diffraction data), the atomic-scale structural interactions between water, urea and the salt (TFA) with KGPGK have been assessed.

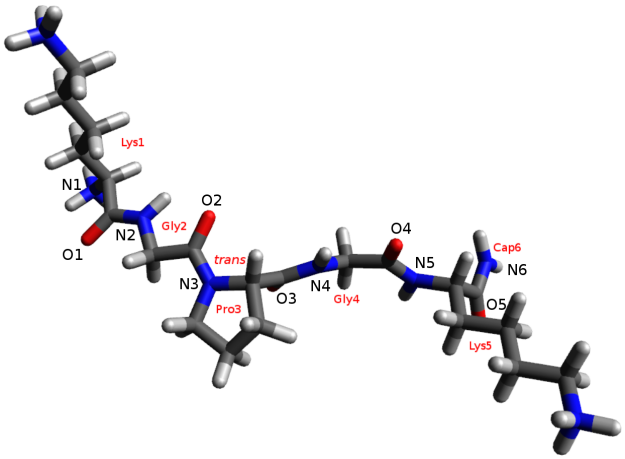

Figure 1 Molecular structure of the $^{+}H_3N$-KGPGK-NH_2 (*trans*) pentapeptide, including the labeling scheme used in this study

Materials and methods

Neutron and X-ray diffraction

Lysine-glycine-proline-glycine-lysinamide with trifluoroacetic acid ($KGPGK^{3+}\cdot 3TFA^{-}$) was purchased from Proteogenix and its purity was verified by elemental analysis (MEDAC Ltd) and NMR (see SI). Urea was purchased from Sigma Aldrich and used without further purification. For the deuterium containing samples the exchangeable hydrogens of the peptide and urea were substituted with deuterium by dissolving the peptide in D_2O (99.8%) and then freeze drying. This process was repeated three times in order to ensure sufficient deuteration. The samples for neutron and X-ray diffraction measurements were prepared by weight with the appropriately labelled peptide, H_2O and/or D_2O (99.98 %) to a molecular ratio of 1:346 (peptide:water) and 1:2.75:346 (peptide:urea:water). The measured pH/pD of KGPGK:TFA in water was 2 and in the KGPGK:urea:TFA the pH/pD was 4.

Neutron diffraction data for five isotopically substituted samples (see Table 1) of KGPGK

in solution, both with and without urea, were collected on the SANDALS diffractometer at the ISIS Facility (STFC, UK) for 8 hours per sample at standard temperature and pressure. The ISIS neutron producing target (target station 1) is driven by a 50 Hz, 800 MeV, 200 μA proton beam from a rapid cycling synchrotron.

Table 1 Diffraction samples; The level of isotopic substitution refers to the exangeable hydrogens

Sample	Isotopic substitution
Neutron 1	100% H
Neutron 2	64% H, 36% D
Neutron 3	50% H, 50% D
Neutron 4	36% H, 64% D
Neutron 5	100% D
X-ray 6	100% H

Data were also collected for the empty cells, the empty instrument and a vanadium standard for background subtraction and normalization of the measured diffraction data for KGPGK solutions. The samples were measured in Hellma quartz containers with a sample thickness of 1 mm and a wall thickness of 1 mm equipped with JYoung's valves. For each data set, the measured scattering levels were within 10% of the expected theoretical values.[28] The data were corrected for absorption, multiple scattering and inelasticity effects, normalized to a vanadium standard and converted to the structure factor $F(Q)$ using the GudrunN program.[29] X-ray diffraction data were collected for the fully protiated sample contained in 0.1 mm diameter borosilicate glass capillary using a Panalytical X'pert Pro diffractometer, also available at the ISIS facility, which has an incident wavelength of 2 Å. The measured data were reduced and corrected for effective density, absorption, bremsstrahlung and Compton scattering and normalized to the total form factor and converted to a structure factor using GudrunX.[29]

The structure factor is a measure of the atomic structure in solution and is the sum of partial structure factors $S_{\alpha\beta}(Q)$ for each unique atomic component in solution and can be

written as:

$$F(Q) = \sum_{\alpha,\beta \geq \alpha} (2 - \delta_{\alpha\beta}) c_\alpha \, c_\beta \, b_\alpha \, b_\beta \, (S_{\alpha\beta}(Q) - 1) \tag{1}$$

where c_i and b_i are the relative concentration and scattering length of atom i[28] (α or β), $\delta_{\alpha\beta}$ is the Kronecker delta function, Q is the scattering vector, $Q = 4\,\pi/\lambda \cdot \sin(2\theta/2)$ with the neutron wavelength λ and the scattering angle 2θ. The corresponding Fourier transformations of the measured $F(Q)$ data, the sum of radial distribution functions $G(r)$, are shown in the SI.

The partial structure factors, which are a measure of the atomic distances in reciprocal space can be Fourier transformed which gives radial distribution functions ($g_{\alpha\beta}(r)$; RDFs) in real space via:

$$S_{\alpha\beta}(Q) = 1 + \frac{4\,\pi\,\rho}{Q} \int r \cdot (g_{\alpha\beta}(r) - 1) \cdot \sin(Qr)\,\mathrm{d}r \tag{2}$$

$F(Q)$ can also be obtained using X-rays as a probe, where in this instance the partial structure factors are weighted with respect to their concentration and their coherent atomic form factors $f(Q)$.[30]

Computation

Empirical Potential Structure Refinement

Empirical potential structure refinement (EPSR) is a reverse Monte Carlo simulation technique which fits an atomistic model to a set of experimentally measured X-ray and neutron diffraction data.[27] EPSR uses a box of molecules at the same concentration and density as the experimental diffraction measurements. Starting potentials (see SI) for each unique atom are refined iteratively until the EPSR-simulated diffraction data shows agreement with the experimental data and has been used to determine the atomic scale structure in a variety of solutions.[20,31–38] It should be noted that EPSR, similar to MD, does not provide the only possible model for the solution in question however, it does provide a model which is consistent with a set of measured data, as it is constrained by that data unlike MD, which is

physically reasonable. In the present case, both EPSR simulations (see below) have been constrained to fit six unique data sets, five obtained from neutron diffraction on different isotopomers and one X-ray diffraction pattern.

For the work presented here, the modeling box contained 20 KGPGK molecules (atomic labeling for the $KGPGK^{+3}$ molecules is shown in Fig. 1) , 60 TFA^- ions and 6920 water molecules at the measured density ($\rho = 0.10064$ atoms / $Å^{-3}$) at a temperature of 298 K. For the simulations with urea, 55 molecules of urea were added to this mixture ($\rho = 0.10075$ atoms / $Å^{-3}$). Starting potentials for water molecules were taken from the TIP3P water model[39] and parameters for the peptide molecules and TFA ions from the CHARMM forcefield[40–44] and modified in order to adjust for different atomic labeling in EPSR compared with MD (see SI). The EPSR simulation contained a mixture of *cis* and *trans* KGPGK molecules (with respect to the Gly-Pro bond) in ratios which correspond with that measured by ^{1}H NMR (15% cis; 85% trans; see SI).

Molecular dynamics simulations

MD simulations were performed at the same molecular ratios as the diffraction measurements and EPSR simulations. Each system contained 20 KGPGK-NH_2 molecules with trans peptide bonds which were observed to be the significant majority from the NMR experiments, 60 TFA^- counter ions, 55 urea molecules and 6920 water molecules. The KGPGK-NH_2 molecules were modeled using the CHARMM36 forcefield[40–42] and the TFA^- ions and urea molecules were modeled using the CHARMM General force field.[43,44] The water molecules were modeled with the TIP3P model[39] which has been modified for the CHARMM force field.[45] All of the bonds and angles for the water molecules were constrained using the SHAKE algorithm[46] and the simulations were conducted using GROMACS 4.[47]

The initial configuration for the simulation was constructed using the Packmol software.[48] The first stage of the simulation protocol used in this study is an energy minimisation simulation in order to eliminate any atomic overlaps that resulted from the construction of

the initial configuration. Then a 2 ns simulation utilising the NVT ensemble with a target temperature of 300 K was performed in order to equilibrate the temperature of the system. Then a 2 ns NPT simulation was performed with a target temperature of 300 K and a target pressure of 1 atm, in order to equilibrate the pressure and the volume of the simulated systems. Finally, a NPT production simulation was performed at 300 K and 1 atm for 80 ns with a timestep of 2 fs. The Nose-Hoover thermostat[49,50] was used in all simulations to control the temperature, while the Martyna-Tuckerman-Tobias-Klein (MTTK) barostat[51] was used in the NPT simulations to control the pressure. A cut-off of 14 Å was used for the van der Waals interactions, and the long range Coulomb interactions were calculated using the particle mesh Ewald (PME) algorithm.[52,53]

ANGULA

From the EPSR simulation, structural information such as the average coordination of each radial distribution function can be easily obtained from integration of the individual RDF curves via:

$$n_{\alpha}^{\beta}(r) = 4\pi\,\rho\,c_{\beta}\int_{r_{min}}^{r_{max}} r^2\,g_{\alpha\beta}(r)\,\mathrm{d}r \tag{3}$$

The coordination numbers in Table 3 and 4 were taken between $r_{min} = 0$ Å and the tabulated r_{max} value in Å.

In addition, information about the three dimensional arrangement of molecules, or parts of the molecules, relative to one another can be extracted from EPSR and MD simulation boxes using the program ANGULA.[54,55] ANGULA requires the definition of a coordinate system for each molecule. Figure 2 shows the orthonormal coordinate systems assigned to the KGPGK peptide nitrogens and the water, TFA and urea molecules for the analysis presented here. Using these coordinate systems and 3000 snapshots of both the EPSR and MD simulation boxes, the position of the first neighbor molecules or molecules within a given

distance range of each fragment can be depicted as spatial density maps (SDMs) that show the probability density of one molecule around another in three dimensions.[55,56]

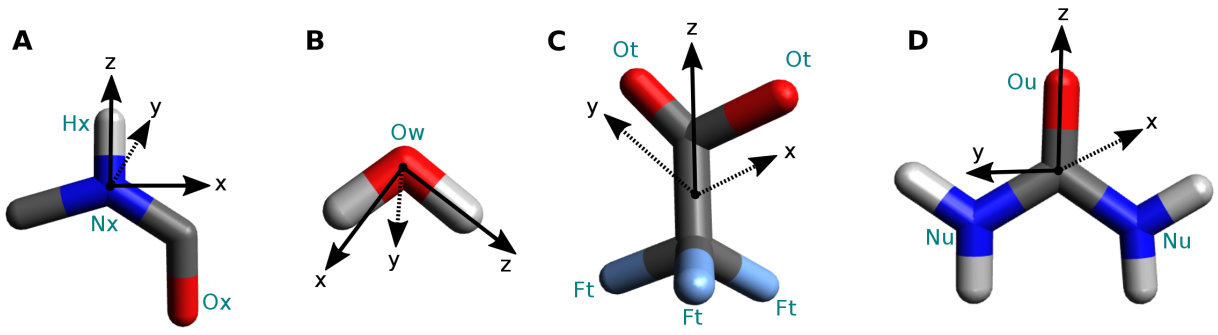

Figure 2 Coordinate systems used in ANGULA assigned to A: the peptide nitrogens N2, N4 and N5; B: water; C: TFA; D: urea.

Results

Diffraction data and EPSR fits

Figure 3 shows the collected neutron and X-ray diffraction data for KGPGK solution samples with and without urea in comparison with the simulated $F(Q)$ fits from EPSR and a comparison with the MD trajectories converted into reciprocal space and the Fourier transformations for these functions in real space are shown in the SI. As expected EPSR fits the neutron and X-ray data very well except for very low Q values in the neutron diffraction data, where discrepancies in this region are likely due to under-estimated inelastic scattering corrections.[57] In MD, $F(Q)$ oscillation at medium Q (6-10 Å^{-1}) are often different from the experimental data which predominantly arise from the TIP3P water model as it has a fixed intramolecular O-H bond distance of 0.95 Å, which is shorter than the measured OH position in the diffraction patterns.

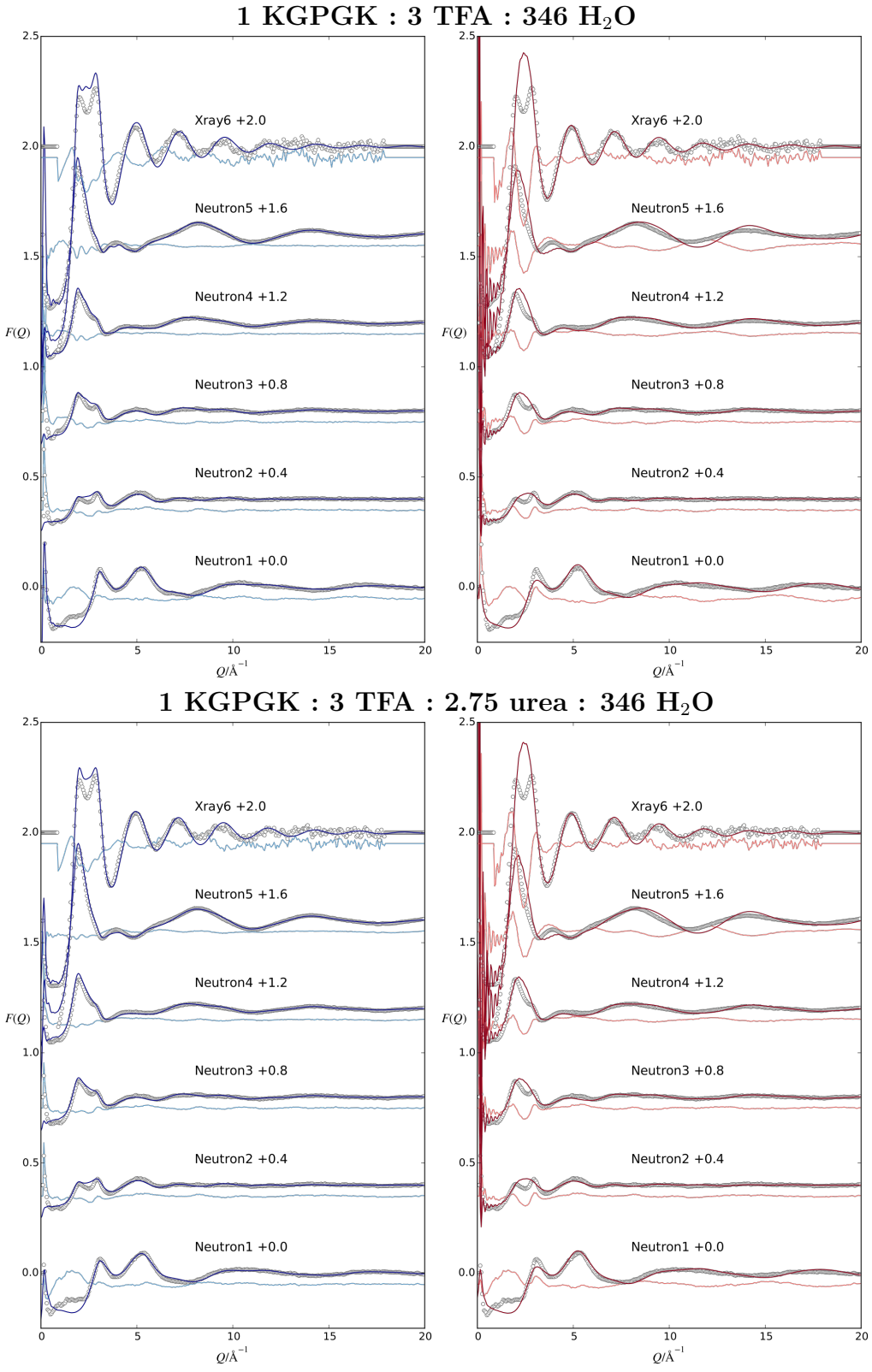

Figure 3 Neutron data (circles) compared EPSR simulated data (left; dark blue line) and the difference between the data and fits (light blue line). Right: MD simulated data (dark red line) and the difference between MD and the measured data (light red line). The level of deuteration is in accordance with the diffraction sample labels in Table 1. The data have been shifted by increments of 0.4 for clarity.

Urea lowers interaction between TFA and peptide backbone

Fig. 4 shows the RDFs for the peptide backbone nitrogen interactions with the water oxygens (Ow), TFA (Ot) and urea (Ou). The corresponding coordination numbers of the first RDF-peaks can be found in Table 2 (a complete overview of all backbone nitrogen interactions are included in the SI). From this figure, the two simulation techniques (EPSR and MD) show very similar results for the hydration of the peptide nitrogens (N1, N2, N4, N5; top row) with the water oxygens (Ow). The positively charged N1 is, relatively, the most hydrated with a distinct peak at 2.7 Å (CN $\sim$ 4.1-4.4). The NH backbone nitrogens N2, N4 and N5 have a smaller hydration peak at 2.8 Å (CN $\sim$ 0.8-0.9). In MD the hydration is almost unchanged and only slightly lowered by the addition of urea (CN $\sim$ -3 %) while in EPSR there is a small increase in nitrogen hydration upon urea addition (CN $\sim$ + 8 %).

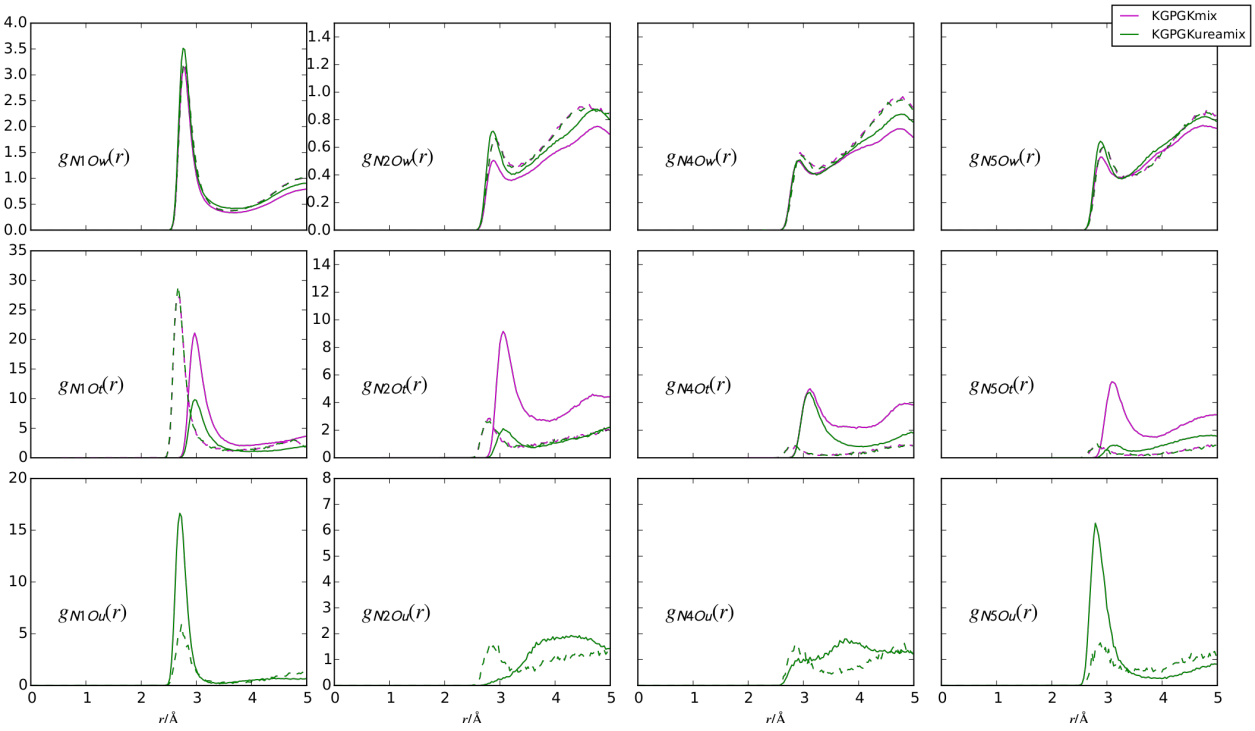

Figure 4 Radial distribution functions between the KGPGK backbone nitrogens (N1-N2, N4-N5) and the solvent and solute molecules (row 1: water oxygen Ow, row 2: TFA oxygen Ot, row 3: urea oxygen Ou). Urea-containing simulations are shown in green, non-urea containing simulations are shown in magenta. The EPSR $g(r)$s are shown as solid lines, the MD $g(r)$s as dashed lines. The corresponding coordination numbers for these functions are in Table 2.

Unlike the peptide hydration, the peptide-TFA interactions (Fig. 4, second row) vary significantly between EPSR and MD. In the MD simulations, the N-terminus (N1) is preferentially coordinated by TFA compared with the polar NH groups that show a much smaller interaction with this anion. There is no influence of urea visible in MD as the coordination numbers are unchanged by the presence of urea (Table 2). In EPSR in the other hand, the TFA-oxygens interact much less (- 30-80 %) with the peptide backbone nitrogens compared with MD upon the addition of urea to the solution; the first peak in the N1-Ot RDF at $\sim$ 3 Å is further away from the nitrogen compared with the same RDF in MD ($\sim$ 2.7 Å). In general, the interactions with the charged N1 N-terminus is lower for EPSR compared to MD while TFA interaction with the NH backbone nitrogens is significantly higher in the EPSR simulations.

Table 2 Coordination numbers for the RDFs in Fig. 4

		KGPGK		KGPGK+UREA	
	$\mathbf{r}_{max}$ / **Å**	EPSR	MD	EPSR	MD
N1-Ow	**3.72**	4.14	4.31	4.40	4.23
N2-Ow	**3.21**	0.76	0.96	0.93	0.92
N4-Ow	**3.21**	0.78	0.81	0.78	0.77
N5-Ow	**3.21**	0.80	0.85	0.84	0.84
N1-Ot	**3.84**	0.58	0.48	0.27	0.48
N2-Ot	**3.84**	0.36	0.11	0.08	0.11
N4-Ot	**3.84**	0.24	0.03	0.17	0.03
N5-Ot	**3.84**	0.22	0.03	0.05	0.03
N1-Ou	**3.36**	-	-	0.10	0.04
N2-Ou	**3.51**	-	-	0.01	0.02
N4-Ou	**3.51**	-	-	0.02	0.02
N5-Ou	**3.90**	-	-	0.06	0.03

The urea-oxygen $g(r)$s (Fig. 4, third row) are again different for EPSR and MD. For EPSR, the urea interaction (Ou) varies widely between the backbone nitrogen sites with N1 showing the highest preference for interacting with the urea, N2 and N3 showing an absence of hydrogen bonding to urea and N5 showing a strong hydrogen bonding interaction with Ou. In MD, urea interacts much less with the peptide nitrogens in general compared with EPSR, showing a preference to bind to the charged N-terminus rather than any of the backbone -NH

groups. However, it is worth noting that the urea-backbone nitrogen coordination numbers are relatively small (Table 2) for both EPSR and MD when compared to TFA or water coordination to the peptide.

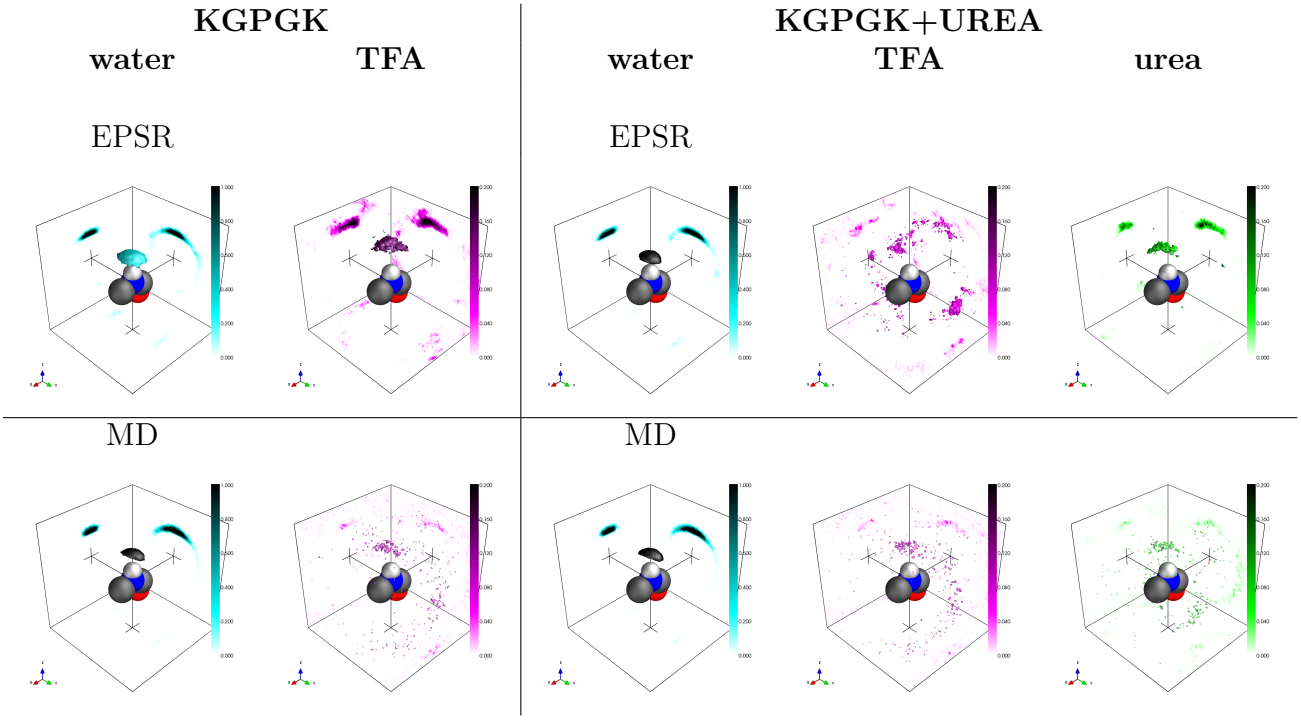

Figure 5 Spatial density maps (SDMs) show the location of the first neighbor water (light blue), TFA (magenta) and urea (green) molecules around the peptide backbone nitrogen N5 for the EPSR (top row) and MD (bottom row) simulations. The isocontour surfaces enclose the most dense 50% for water and 10% for TFA and urea. The scale bar show the local density of neighbors around the central nitrogen in units of atoms/Å^3.

The SDMs of the Gly-Lys N5 nitrogen with solvent and solute molecules in Fig. 5 exemplify the three-dimensional arrangement of nearest neighbor water (light blue), TFA (magenta) and urea (green) molecules around the peptide backbone nitrogens (a complete overview of the other peptide backbone nitrogen SDMs are in the SI). The position of hydrating water molecules is similar in both simulations. The water becomes slightly more dense and localized directly above the hydrogen atom in EPSR upon the addition of urea. Overall, TFA is localized in similar positions to water, as evidenced by Fig. 5, but the localization and density of TFA is significantly higher in EPSR compared with MD. Similar to the RDFs in Fig. 4, the lower TFA density in MD is unchanged by the addition of urea. However, in EPSR the

addition of urea disrupts the highly localized and directed TFA-nitrogen interactions as the TFA density cloud directly above N5 completely disappears and becomes scattered around the nitrogen similar to the accordant SDM in MD. This is consistent with the replacement of TFA with urea molecules as shown in the RDFs for N5 in Fig. 4. It should be noted that the urea density, although higher in EPSR than in MD, does not exactly replace the TFA density, in either simulation. Therefore, urea likely only partially replaces TFA but also acts to increase the hydration around the N5 site. The SDMs for the other nitrogens (SI), show that both N1 and N2 also show decreased interactions with TFA upon urea addition in EPSR, similar to N5.

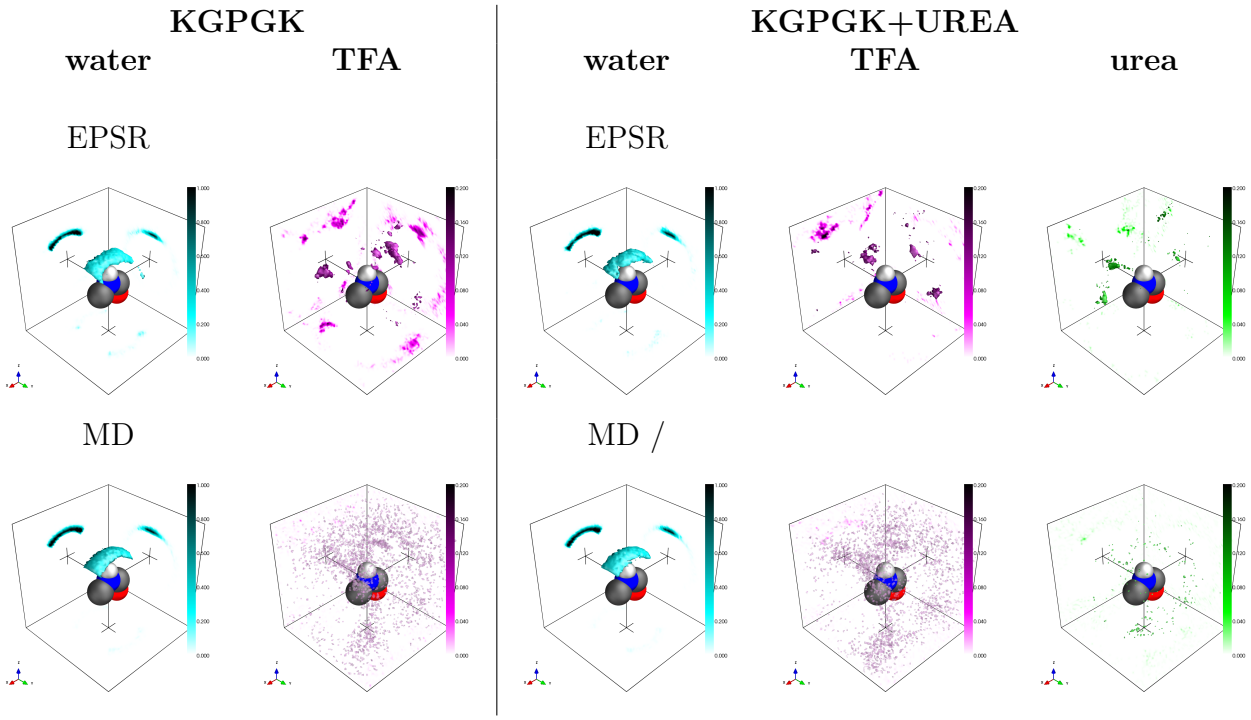

Figure 6 Spatial density maps (SDMs) show the location of the first neighbor water (light blue), TFA (magenta) and urea (green) molecules around the peptide backbone nitrogen N4 for the EPSR (top row) and MD (bottom row) simulations. The isocontour surfaces enclose the most dense 50% for water and 10% for TFA and urea. The scale bar show the local density of neighbors around the central nitrogen in units of atoms/Å^3.

As a comparison, Fig. 6 shows the influence of urea on the Pro-Gly N4 nitrogen. Comparable to N5, the water density directly above the NH is almost identical between both simulation methods and unchanged by the presence of urea, where the respective coordina-

tion numbers (N4-Ow) are also identical (~ 0.8; Table 2). Similar to N5, TFA interacts more with N4 in the EPSR simulations than in the MD simulations, however this interaction is much less localized compared to the TFA interaction with N5 in EPSR (see Fig. 5) and consequentially the TFA-N4 interaction is less disrupted by the addition of urea. The amide cap terminal nitrogen (N6) shows a TFA interaction that is almost unchanged by urea, similar to N4 (see SI).

Overall, these results show a comparable hydration of the peptide backbone in MD and EPSR both in the presence and absence of urea. On the other hand, the MD simulations only show limited interactions between the solutes TFA and urea with the peptide backbone, whereas by contrast in the EPSR, urea disrupts the strong TFA-peptide interaction and increases peptide hydration slightly in general.

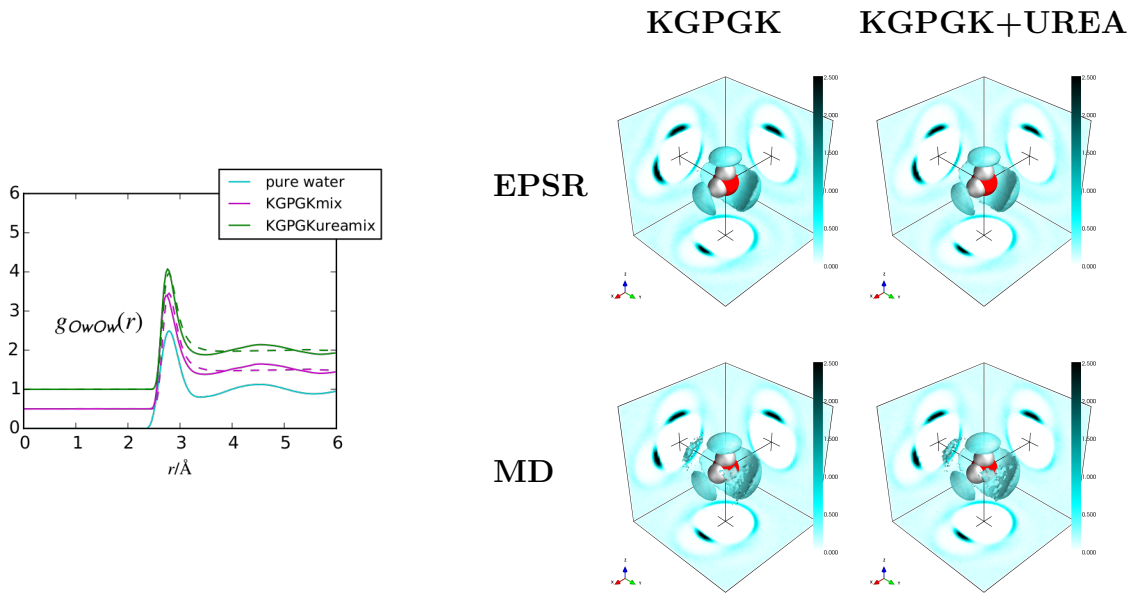

Figure 7 Left: Radial distribution function ($g(r)$s) between the water oxygens (Ow) with (urea) and without urea (magenta) compared with those from pure water.[58] The EPSR $g(r)$s are shown as solid lines and the MD $g(r)$s as dashed lines. The corresponding coordination numbers are in Table 3. Right: Spatial density maps (SDMs) show the location of the water (light blue) molecules up to a 10 Å distance around a central representative water molecule for the EPSR (top row) and MD (bottom row) simulations. The isocontour surfaces enclose the most dense 5% for water. The scale bar show the local density of neighbors around the central oxygen in units of atoms/Å^3.

Urea does not change the bulk water structure

To determine whether urea changes not only the peptide interactions with both water and TFA but also changes the bulk water-water structure, as is proposed in the theory of indirect urea-induced protein denaturation,[3] Fig. 7 (left) shows the water (Ow-Ow) RDF of both simulation techniques compared to the Ow-Ow $g(r)$ from EPSR simulations of pure water.[58] From this figure, it is clear that neither the nearest neighbor water-water nor the second nearest neighbor water-water structure is significantly changed from that of pure water, with the exception of the second nearest neighbor peak in the MD, as expected from the TIP3P water model from MD simulations. From the coordination numbers in Table 3, EPSR shows a slight increase in water-water interaction with KGPGK addition and a small decrease with urea addition, while MD shows a slight increase in water-water interaction upon the addition of urea. The water-water SDMs in Figure 7 (right) show no difference to the tetrahedral water-water structure upon urea addition to the solution in either simulation, however, again as expected the MD simulations show a slightly more diffuse second nearest neighbor water coordination due to the TIP3P model.

Table 3 Coordination numbers for the RDFs in Fig. 7, 8 and 9

		pure water[58]	KGPGK		KGPGK+UREA	
	$\mathbf{r}_{max}$ / **Å**	EPSR	EPSR	MD	EPSR	MD
Ow-Ow	**3.48**	5.15	5.32	5.30	4.89	5.24
Ot-Ow	**3.96**		4.80	5.31	4.83	5.26
Ft-Ft	**3.93**		0.54	0.11	0.24	0.11

Urea does not change the TFA-hydration

One of the largest differences between the simulation methods are the resulting TFA-peptide interactions, which is, perhaps, due to the change in TFA-TFA and TFA-water association upon the addition of urea between the simulations. Figure 8 (left) shows the $g(r)$ between TFA oxygens (Ot) and water oxygens (Ow) and the corresponding spatial density maps in Fig. 8 (right) show the location of nearest neighbor water molecules around the whole

TFA molecule. In the MD simulations, the negatively charged TFA molecules interact more with water compared with the EPSR fits to the neutron diffraction data, where the most localized hydration is visible around the negatively charged TFA oxygens. The nearest neighbor hydration around TFA is denser and more localized for MD compared to EPSR, both in terms of the three-dimensional arrangement and the radial distance from the TFA oxygens (Fig. 8, left). It should be noted that the first MD $g(r)$ peak is at 2.7 Å, closer to the TFA oxygen than in EPSR ($\approx$ 3.0 Å). Similar to peptide N-terminus N1 hydration, both simulations show no significant difference in TFA hydration for the KGPGK mixture with urea (see Table 3). However, there is also a lack of increase in hydration upon the addition of urea which might be expected if urea were to replace interactions of TFA either with itself (see below) or with the peptide backbone. This is likely a result of the fact that urea forms direct hydrogen bonds to TFA (see SI), explaining the lack of additional hydration around TFA in solution.

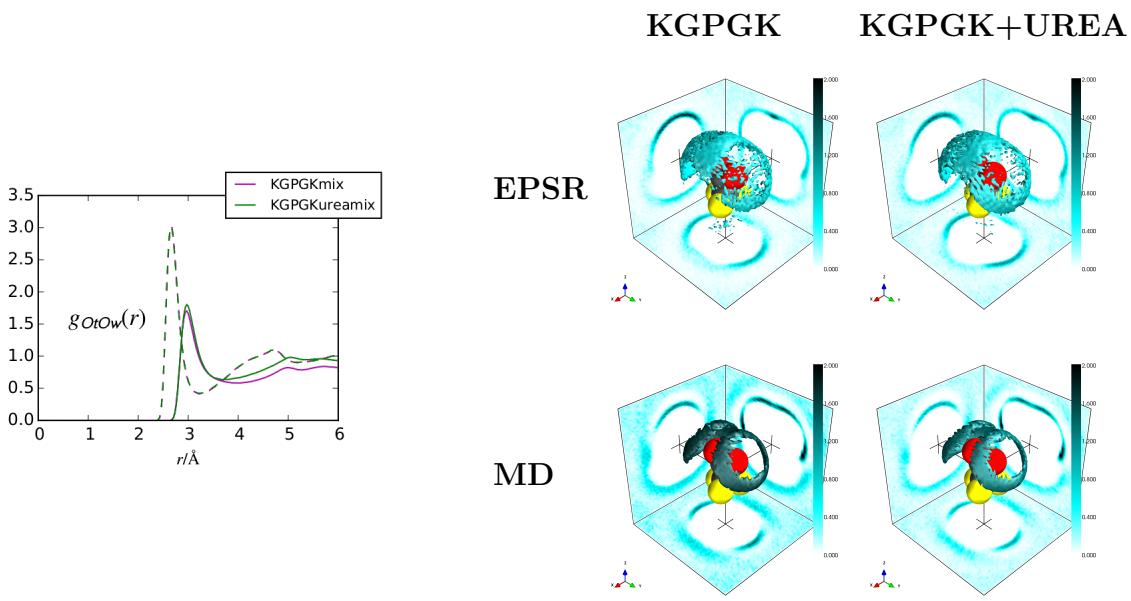

Figure 8 Left: Radial distribution function ($g(r)$s) between the TFA oxygen (Ot) and water oxygens (Ow)with (green) and without urea (magenta) for EPSR (solid lines) and MD (dashed lines) simulations. The corresponding coordination numbers are in Table 3 Right: Spatial density maps (SDMs) show the location of the water (light blue) molecules up to a 10 Å distance around the middle of a central representative TFA molecule for the EPSR (top row) and MD (bottom row) simulations. The isocontour surfaces enclose the most dense 5% for water. The scale bar show the local density of neighbors around the central oxygen in units of atoms/Å^3.

Urea decreases TFA-TFA-clustering

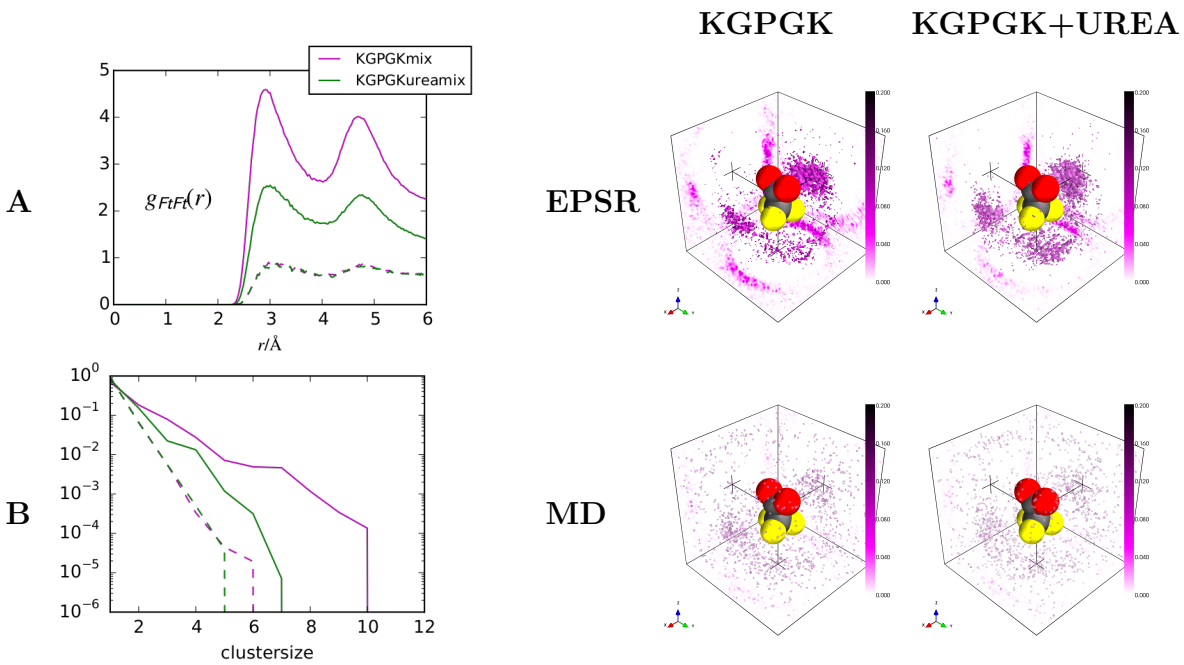

Figure 9 Left: A) Intermolecular $g(r)$s (EPSR solid lines, MD dashed lines) between the TFA fluorines (Ft). B) Proportion of TFA-TFA association within a 0-4 Å intermolecular Ft-Ft distance.
Urea-containing simulations are shown in green, non-urea-containing simulations are shown in magenta. The corresponding coordination numbers are in Table 3. Right: Spatial density maps (SDMs) show the location of the TFA (magenta) molecules up to a 10 Å distance around a central representative TFA molecule for the EPSR (top row) and MD (bottom row) simulations. The isocontour surfaces enclose the most dense 10% for water. The scale bar show the local density of neighbors around the central TFA molecules in units of atoms/Å^{-3}.

A possible explanation for the lack of TFA hydration in EPSR compared with MD is the higher degree of TFA-TFA association in EPSR. In order to probe this, Fig. 9 shows the $g(r)$s for the TFA fluorine atoms with other TFA molecules in solution with the coordination numbers for these functions shown in Table 3. (Other intermolecular TFA-TFA $g(r)$s are shown in the SI.) Not only does EPSR show a greater number of TFA-TFA interactions, these interactions are markedly lowered upon the addition of urea, unlike in MD where the TFA-TFA $g(r)$ is unchanged by urea addition. From the SDMs in Fig. 9, TFA-TFA association in EPSR mostly occurs via the fluorine and carbon atoms as visible in the high TFA density next to the TFA carbon atoms. It should be noted that TFA-TFA interactions

are heavily weighted in the X-ray data and EPSR is in better agreement with this data set compared with the MD $F(Q)$ data.

Fig. 9 also shows the results of a clustering analysis between TFA molecules in solution for both MD and EPSR simulations. A cluster was defined as two TFA molecules within a 0-4 Å intermolecular Ft-Ft distance. Most TFA molecules are non-associated with each other but the small proportion of larger TFA-TFA clusters is mostly present in the EPSR simulations. The addition of urea significantly decreases the relative cluster size and the proportion of TFA molecules in clusters.

Overall these results show a significant difference between EPSR and MD regarding the TFA interactions. MD underestimates TFA-TFA association (Fig. 9) and TFA-peptide interaction (Fig. 4) and overestimates TFA-hydration (Fig. 8).

Conclusions

In the current work the hydration of the KGPGK peptide is either mostly unchanged (for MD) or slightly increased (for EPSR) upon urea addition to the solution. This is contrary to what occurs for the small peptide, GPG, in aqueous solution with Cl^- as the counter ion, where the presence of urea (3.3 M) lowers the peptide backbone hydration by up to -20 %.[7] A similar replacement of water by urea has also been observed in X-ray crystallography experiments on lysozyme.[59] The relative absence of water replacement around the KGPGK backbone upon the addition of urea in the present work, suggests that the presence of TFA in the solutions largely prevents the association of urea with the peptide, perhaps, effectively protecting the peptide from denaturation. Similarly, the addition of urea to the solution does not largely disrupt either the bulk water structure or the TFA hydration in the solutions, where only minor changes in the coordination numbers are observed in either simulation.

In the absence of urea, MD shows very few peptide-TFA interactions along the peptide backbone. These interactions are more prevalent in the EPSR simulation and vary, along

the backbone. Upon the addition of urea to the solution, these peptide-TFA interactions decrease in EPSR and remain roughly the same in the MD simulation, presumably because there were relatively few interactions present in this simulation to begin with. In EPSR, urea disrupts the TFA molecules interacting with the peptide, in some cases by replacing these interactions but only to a very minor extent, given the very low coordination numbers for peptide-urea interactions (Table 2). This difference between EPSR and MD simulations appears to be a result of the TFA being less hydrated and associating more with itself in EPSR compared to MD.

Interestingly, overall the TFA interactions with either the peptide or between TFA molecules is decreased upon the addition of urea to the solution in the EPSR simulation, while these interactions remain relatively constant in the presence and absence of urea in MD. The discrepancies between the two simulation methods may be due to the fact MD does not as accurately predict the heavy atom positions for TFA as EPSR does in this particular solution. This is evidenced by the fact that the largest differences between the two simulations compared with the measured diffraction data in Fig. 3 occurs in the X-ray diffraction pattern, where the heavy atom positions of TFA are more heavily weighted, and it is clear from this figure that the EPSR model provides a better fit to this data set than MD and as such the EPSR likely provides a more accurate picture of the TFA interactions in the solution.

The addition of urea leads to a decrease in charged-charged interaction in the solution - both between the polar peptide bonds in KGPGK and TFA and also between TFA molecules with themselves. This, perhaps, suggests that the ability of urea to denature proteins is dependent both upon the pH and ionic strength of the solution. It may be that urea, at low pH and high ionic strength (as in this study), first disrupts the peptide-salt interactions and only at higher urea concentrations directly interacts with the peptide, replacing the hydrating water molecules, as has been previously observed in the absence of high salt concentration at a higher pH.[5–7,9,10] Urea-small molecule interactions have been suggested to contribute to

the osmolytic activity of betaine,[15] where it was found that betaine effectively removed urea from the surface of a protein preventing its subsequent denaturation.[17]

Further, the fact that high salt concentrations can protect against denaturation by urea is in accordance with previous studies showing the protective influence of higher salt concentrations[14] against urea-induced unfolding on human serum albumin protein. MD simulations on amino acid side chain analogues and protein backbone mimic molecules with urea and NaCl have suggested that the high salt protection against unfolding of human serum albumin is due to an opposing salting effect between urea and the salt.[60] This hypothesis would also explain protective properties of high-salt buffers for proteins or the salt resistance of certain organisms.[61]

The emerging picture from this work is that at high salt concentrations, the salt-urea interactions can help to protect peptides and proteins from denaturation by intercepting the urea molecules by direct salt-urea contacts. It may be that at even higher salt concentrations, that salts can also effectively block access to the peptide-charged sites as there is a small propensity for the salt in the present solution, to interact with the polar portions of the peptide backbone even at the current salt concentrations. Similarly, it may also be that at much higher urea concentrations, salt-urea interactions will not be sufficient to prevent denaturation.

Acknowledgements

We thank the ISIS Facility (Rutherford Appleton Laboratories, STFC, UK) for the allocation of neutron beam time and the UK Engineering and Physical Sciences Research Council (EP/J002615/1) and the Leverhulme Trust (RPG-2015-135) for funding. The MD simulations were performed via our membership of the UKs HEC Materials Chemistry Consortium, which is funded by EPSRC (EP/L000202), this work used the ARCHER UK National Supercomputing Service (http://www.archer.ac.uk). Additionally, CDL and AG thank the

Faculty of Natural and Mathematical Sciences and the Department of Physics at King's College London for funding AG's summer internship.

References

(1) Dill, K. A.; Shortle, D. Denatured states of proteins. *Annual review of biochemistry* **1991**, *60*, 795–825.

(2) Anfinsen, C. B. Principles that Govern the Folding of Protein Chains. *Science* **1973**, *181*, 223–230.

(3) Rezus, Y.; Bakker, H. Effect of urea on the structural dynamics of water. *Proceedings of the National Academy of Sciences* **2006**, *103*, 18417–18420.

(4) Canchi, D. R.; García, A. E. Backbone and side-chain contributions in protein denaturation by urea. *Biophys. J.* **2011**, *100*, 1526–1533.

(5) Moeser, B.; Horinek, D. Unified description of urea denaturation: backbone and side chains contribute equally in the transfer model. *The Journal of Physical Chemistry B* **2013**, *118*, 107–114.

(6) Su, Z.; Dias, C. L. Molecular interactions accounting for protein denaturation by urea. *Journal of Molecular Liquids* **2016**,

(7) Steinke, N.; Gillams, R. J.; Pardo, L. C.; Lorenz, C. D.; McLain, S. E. Atomic scale insights into urea–peptide interactions in solution. *Phys. Chem. Chem. Phys.* **2016**, *18*, 3862–3870.

(8) O'Brien, E. P.; Dima, R. I.; Brooks, B.; Thirumalai, D. Interactions between hydrophobic and ionic solutes in aqueous guanidinium chloride and urea solutions: lessons for protein denaturation mechanism. *J. Am. Chem. Soc.* **2007**, *129*, 7346–7353.

(9) Auton, M.; Holthauzen, L. M. F.; Bolen, D. W. Anatomy of energetic changes accompanying urea-induced protein denaturation. *Proceedings of the National Academy of Sciences* **2007**, *104*, 15317–15322.

(10) Stumpe, M. C.; Grubmüller, H. Interaction of urea with amino acids: implications for urea-induced protein denaturation. *Journal of the American Chemical Society* **2007**, *129*, 16126–16131.

(11) Stumpe, M. C.; Grubmüller, H. Urea impedes the hydrophobic collapse of partially unfolded proteins. *Biophys. J.* **2009**, *96*, 3744–3752.

(12) Daidone, I.; Ulmschneider, M. B.; Di Nola, A.; Amadei, A.; Smith, J. C. Dehydration-driven solvent exposure of hydrophobic surfaces as a driving force in peptide folding. *Proc. Nat. Ac. Sci.* **2007**, *104*, 15230–15235.

(13) Ahmad, F.; Bigelow, C. C. Inorganic salt denaturants stabilize ribonuclease against denaturation by urea. *Canadian journal of biochemistry* **1978**, *56*, 1003–1005.

(14) Muzammil, S.; Kumar, Y.; Tayyab, S. Anion-induced stabilization of human serum albumin prevents the formation of intermediate during urea denaturation. *Proteins: Structure, Function, and Bioinformatics* **2000**, *40*, 29–38.

(15) Wang, A.; Bolen, D. A naturally occurring protective system in urea-rich cells: mechanism of osmolyte protection of proteins against urea denaturation. *Biochemistry* **1997**, *36*, 9101–9108.

(16) Dodia, M.; Bhimani, H.; Rawal, C.; Joshi, R.; Singh, S. Salt dependent resistance against chemical denaturation of alkaline protease from a newly isolated haloalkaliphilic Bacillus sp. *Bioresource technology* **2008**, *99*, 6223–6227.

(17) Chen, J.; Gong, X.; Zeng, C.; Wang, Y.; Zhang, G. Mechanic Insight into Resistance

of Betaine to Urea Induced Protein Denaturation. *The Journal of Physical Chemistry B* **2016**,

(18) Pace, C. N.; Laurents, D. V.; Erickson, R. E. Urea denaturation of barnase: pH dependence and characterization of the unfolded state. *Biochemistry* **1992**, *31*, 2728–2734.

(19) Pace, C. N.; Laurents, D. V.; Thomson, J. A. pH dependence of the urea and guanidine hydrochloride denaturation of ribonuclease A and ribonuclease T1. *Biochemistry* **1990**, *29*, 2564–2572.

(20) Busch, S.; Bruce, C. D.; Redfield, C.; Lorenz, C. D.; McLain, S. E. Water Mediation Is Essential to Nucleation of β-Turn Formation in Peptide Folding Motifs. *Angew. Chem. - Int. Ed.* **2013**, *52*, 13091–13095.

(21) Cheung, M. S.; García, A. E.; Onuchic, J. N. Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proc. Nat. Ac. Sci* **2002**, *99*, 685–690.

(22) Lewis, P. N.; Momany, F. A.; Scheraga, H. A. Folding of polypeptide chains in proteins: a proposed mechanism for folding. *Proc. Nat. Ac. Sci* **1971**, *68*, 2293–2297.

(23) Chou, P. Y.; Fasman, G. D. Conformational parameters for amino acids in helical, β-sheet, and random coil regions calculated from proteins. *Biochem.* **1974**, *13*, 211–222.

(24) Wilmot, C.; Thornton, J. Analysis and prediction of the different types of β-turn in proteins. *J. Mol. Bio.* **1988**, *203*, 221–232.

(25) Guruprasad, K.; Rajkumar, S. Beta-and gamma-turns in proteins revisited: a new set of amino acid turn-type dependent positional preferences and potentials. *J. Biosci.* **2000**, *25*, 143–156.

(26) Website, `http://www.ebi.ac.uk/pdbe-site/pdbemotif/`; 19th October 2016.

(27) Soper, A. K. Computer simulation as a tool for the interpretation of total scattering data from glasses and liquids. *Mol. Sim.* **2012**, *38*, 1171–1185.

(28) Sears, V. F. Neutron scattering lengths and cross sections. *Neutron News* **1992**, *3*, 26–37.

(29) Soper, A. K. *GudrunN and GudrunX: programs for correcting raw neutron and X-ray diffraction data to differential scattering cross section*; Science & Technology Facilities Council, 2011.

(30) Hubbell, J.; Veigele, W.; Briggs, E.; Brown, R.; Cromer, D.; Howerton, R. *J. Phys. Chem. Ref. Data* **1975**, *4*, 471–501.

(31) McCune, J. A.; Turner, A. H.; Coleman, F.; White, C. M.; Callear, S. K.; Youngs, T. G. A.; Swadzba-Kwasny, M.; Holbrey, J. D. Association and liquid structure of pyridine-acetic acid mixtures determined from neutron scattering using a 'free proton' EPSR simulation model. *Phys. Chem. Chem. Phys.* **2015**, *17*, 6767–6777.

(32) Shephard, J. J.; Callear, S. K.; Imberti, S.; Evans, J. S. O.; Salzmann, C. G. Microstructures of negative and positive azeotropes. *Phys. Chem. Chem. Phys.* **2016**, *18*, 19227–19235.

(33) Callear, S. K.; Johnston, A.; McLain, S. E.; Imberti, S. Conformation and interactions of dopamine hydrochloride in solution. *J. Chem. Phys.* **2015**, *142*, 014502.

(34) Johnston, A. J.; Busch, S.; Pardo, L. C.; Callear, S. K.; Biggin, P. C.; McLain, S. E. On the atomic structure of cocaine in solution. *Physical Chemistry Chemical Physics* **2016**, *18*, 991–999.

(35) Sridhar, A.; Johnston, A. J.; Varathan, L.; McLain, S. E.; Biggin, P. C. The solvation structure of Alprazolam. *Physical Chemistry Chemical Physics* **2016**, *18*, 22416–22425.

(36) Tavagnacco, L.; Brady, J. W.; Bruni, F.; Callear, S.; Ricci, M. A.; Saboungi, M. L.; Cesáro, A. Hydration of Caffeine at High Temperature by Neutron Scattering and Simulation Studies. *J. Phys. Chem. B* **0**, *0*, null, PMID: 26421842.

(37) Hayes, R.; Imberti, S.; Warr, G. G.; Atkin, R. The Nature of Hydrogen Bonding in Protic Ionic Liquids. *Angewandte Chemie International Edition* **2013**, *52*, 4623–4627.

(38) Sillrn, P.; Swenson, J.; Mattsson, J.; Bowron, D.; Matic, A. The temperature dependent structure of liquid 1-propanol as studied by neutron diffraction and EPSR simulations. *The Journal of Chemical Physics* **2013**, *138*.

(39) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(40) MacKerell Jr, A. D.; Bashford, D.; Bellott, M.; Dunbrack Jr, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S. et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(41) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E.; Mittal, J.; Feig, M.; MacKerell Jr, A. D. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain $\chi1$ and $\chi2$ dihedral angles. *J. Chem. Theor. Comp.* **2012**, *8*, 3257–3273.

(42) MacKerell, A. D.; Feig, M.; Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comp. Chem.* **2004**, *25*, 1400–1415.

(43) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I. et al. CHARMM general force field: A

force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of computational chemistry* **2010**, *31*, 671–690.

(44) Yu, W.; He, X.; Vanommeslaeghe, K.; MacKerell, A. D. Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations. *Journal of computational chemistry* **2012**, *33*, 2451–2468.

(45) Reiher, W. E. Theoretical studies of hydrogen bonding. Ph.D. thesis, 1985.

(46) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **1977**, *23*, 327–341.

(47) Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of chemical theory and computation* **2008**, *4*, 435–447.

(48) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: a package for building initial configurations for molecular dynamics simulations. *Journal of computational chemistry* **2009**, *30*, 2157–2164.

(49) Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Molecular physics* **1984**, *52*, 255–268.

(50) Hoover, W. G. Canonical dynamics: equilibrium phase-space distributions. *Physical review A* **1985**, *31*, 1695.

(51) Martyna, G. J.; Tuckerman, M. E.; Tobias, D. J.; Klein, M. L. Explicit reversible integrators for extended systems dynamics. *Molecular Physics* **1996**, *87*, 1117–1157.

(52) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N log (N) method for Ewald sums in large systems. *The Journal of chemical physics* **1993**, *98*, 10089–10092.

(53) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *The Journal of chemical physics* **1995**, *103*, 8577–8593.

(54) Website, `https://gcm.upc.edu/en/members/luis-carlos/angula/ANGULA`; September 2014.

(55) Busch, S.; Lorenz, C. D.; Taylor, J.; Pardo, L. C.; McLain, S. E. Short-Range Interactions of Concentrated Proline in Aqueous Solution. *J. Phys. Chem. B* **2014**, *118*, 14267–14277.

(56) Busch, S.; Pardo, L. C.; O'Dell, W. B.; Bruce, C. D.; Lorenz, C. D.; McLain, S. E. On the structure of water and chloride ion interactions with a peptide backbone in solution. *Phys. Chem. Chem. Phys.* **2013**, *15*, 21023–21033.

(57) A.K.Soper, Inelasticity corrections for time-of-flight and fixed wavelength neutron diffraction experiments. *Mol. Phys.* **2009**, *107*, 1667–1684.

(58) Soper, A. K. The Radial Distribution Functions of Water as Derived from Radiation Total Scattering Experiments: Is There Anything We Can Say for Sure? *ISRN Physical Chemistry* **2013**, *2013*, 1–67.

(59) Raskar, T.; Khavnekar, S.; Hosur, M. Time-dependent X-ray diffraction studies on urea/hen egg white lysozyme complexes reveal structural changes that indicate onset of denaturation. *Scientific Reports* **2016**, *6*.

(60) Li, W.; Zhou, R.; Mu, Y. Salting effects on protein components in aqueous NaCl and urea solutions: toward understanding of urea-induced protein denaturation. *The Journal of Physical Chemistry B* **2012**, *116*, 1446–1451.

(61) Yancey, P. H.; Clark, M. E.; Hand, S. C.; Bowlus, R. D.; Somero, G. N. Living with water stress: evolution of osmolyte systems. *Science* **1982**, *217*, 1214–1222.